基于数据内在结构特征的度量学习 *

张开放[1]，郎贵林[1]，王晓明[1,2]，黄增喜[1]，杜亚军[1]

(1. 西华大学计算机与软件工程学院，成都 610039; 2. 西华大学机器人研究中心，成都 610039)

摘　要：半正定约束度量学习(PCML)，作为一种结合了支持向量机(SVM)的典型度量学习方法，在图像识别和行人重识别领域展现了优越的性能。然而，在每次学习度量矩阵的过程中，该方法只简单的考虑不同类别样本之间的最大间隔，忽略了同一类别间的样本特征空间也在发生变化。基于此，提出了一种基于数据内在结构特征的度量学习方法。首先，与 PCML 相比，提出的方法不仅考虑了不同类别样本之间的间隔，而且考虑了相同类别样本间的类内散度矩阵，使学习到的度量矩阵有更强的鉴别能力。其次，进一步将 L1-norm 损失函数转换为 L2-norm 损失函数，这样可以进一步提高模型的泛化性能。最终，在多个数据集上的实验结果表明，多数情况下提出的方法相比于其他度量学习方法取得了更优异的性能。

关键词：度量学习；支持向量机；类内散度；图像识别；行人重识别

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2020.03.0054

Metric learning based on intrinsic structural characteristics of data

Zhang Kaifang[1], Lang Guilin[1], Wang Xiaoming[1,2], Huang Zengxi[1], Du Yajun[1]

(1. *School of Computer & Software Engineer*, *Xihua University*, *Chengdu* 610039, *China*; 2. *Robotics Research Center of Xihua University*, *Chengdu* 610039, *China*)

Abstract: Positive-semidefinite constrained metric learning (PCML) , as a typical metric learning method combined with support vector machine (SVM) , exhibits superior performance in image recognition and person re-identification. However, in the process of learning the metric matrix, this method simply considers the margin between samples of different categories, ignoring that the feature space of samples of the same category also changes. To this end, this paper proposes a metric learning method based on the intrinsic structural characteristics of data. First of all, compared with PCML, the method not only considers the margin between samples of different categories, but also considers the intra-class divergence matrix of samples of the same category, so that the learned metric matrix has stronger discrimination ability. Secondly, this paper further transforms the L1-norm loss function into the L2-norm loss function, which can further improve the generalization performance of the model. Finally, the experimental results on multiple datasets show that the proposed method achieves better performance than other methods in most cases.

Key words: metric learning; support vector machine; intra-class divergence; image recognition; person re-identification

0　引言

过去的十年，度量学习引起了学术界以及工业界越来越多的关注。度量学习或称相似性度量学习是机器学习中一个新兴的研究领域，它旨在通过训练去学习一个有效的度量，使其能够减小同类样本之间的距离，同时增大不同类别样本之间的距离。度量学习在计算机视觉中有着广泛的应用，其中主要包括视觉追踪[1,2]、图像检索[3,4]、人脸识别[5,6]、聚类[7]、行人重识别[8~10]等领域。由于实际的应用中样本的类间相似性和类内差异性，这使得如何获得一个有效的度量成为了当前研究者面临的巨大挑战。

目前度量学习的研究主要集中在马氏距离学习，其主要目的是找到特征空间的全局线性变换，强调相关维度，舍弃不相关维度。由于马氏度量集与多元高斯度量集之间存在双射关系[11]，因此可以用相应的协方差矩阵来表示。此外，马氏距离通过允许特征空间的任意线性缩放和旋转来推广欧氏度量，相比于传统的欧氏度量，马氏距离学习到的度量矩阵具有更强的鉴别能力，在许多实际问题中展现出了优越的性能。马氏距离学习代表性的方法有邻域成分分析(neighborhood components analysis, NCA)[11]、大间隔最近邻(large margin nearest neighbors, LMNN)[12]、信息论度量学习(information theoretic metric learning, ITML)[13]、简单直接的度量学习(keep it simple and straight forward metric learning, KISSME)[14]等。这些方法几乎都是通过利用样本点之间的部分经验知识来优化马氏距离，例如，它们将不同类别中的样本点之间的距离限制为大于相同类中的样本点之间的距离。从理论的角度分析，上面的这些方法都能很好地推广到不可见数据，它们通过修改损失函数或者整合正则化的方案，以避免过拟合现象的发生。

近年来，核学习方法已经广泛的应用到一些学习任务中，例如半监督学习、多实例学习、多任务学习等[15,16]。支持向量机(support vector machine, SVM)[17]，作为典型的核学习方法，在许多实际应用中展现了良好的泛化性能，而且有很多关于核学习方法的开放资源，包括各种工具箱和库已经发布了[18~23]。因此，对于研究者来说，探索如何利用核方法资源来研究和开发新的度量学习方法是非常重要的。最近，一些

收稿日期：2020-03-28；**修回日期：**2020-04-29　**基金项目：**国家自然科学基金资助项目(61602390)；西华大学研究生创新基金项目(ycjj2019085)

作者简介：张开放(1994-)，男，河南周口人，硕士研究生，主要研究方向为机器学习、计算机视觉(1115291605@qq.com)；郎贵林(1995-)，男，四川达州人，硕士研究生，主要研究方向为模式识别、图像处理；王晓明(1977-)，男，四川简阳人，副教授，硕导，博士，主要研究方向为模式识别、机器学习、图像处理、计算机视觉；黄增喜(1985-)，男，广西百色人，副教授，硕导，博士，主要研究方向为深度学习、生物特征识别、图像处理；杜亚军(1967-)，男，四川巴中人，教授，硕导，博士，主要研究方向为数据库理论、智能网络.

研究工作[24, 25]尝试去探索度量学习和 SVM 之间的关系，想利用 SVM 的大间隔原理去开发新的度量学习方法，其中，文献[26]提出一种结合 SVM 的度量学习算法(positive-semidefinite constrained metric learning, PCML)，它将度量学习问题描述为具有半正定约束的核分类问题，通过迭代式训练 SVM 大间隔分类器来指导度量矩阵的学习。重要的是，现有的 SVM 一些求解器，如 LibSVM[16]，可以很方便的应用到求解度量矩阵的过程中。然而，PCML 忽略了这样一个关键事实：在每次通过迭代式训练 SVM 分类器去指导度量矩阵的学习过程中，它只简单的考虑不同类别样本之间的最大间隔，忽略了每一次迭代过程中同一类别间的样本特征空间的也在发生变化，这样限制了模型泛化能力进一步提高。

针对以上不足，本文提出了一种基于数据内在结构特征的度量学习方法。首先，本文提出的方法不仅考虑了不同类别样本之间的间隔，而且考虑了相同类别样本间的类内散度矩阵。其次，参考文献[27]，本文提出的方法进一步将 L_1-norm 损失函数转换为 L_2-norm 损失函数，这样可以进一步提高模型的泛化性能。此外，为了提升模型的计算效率，本文使用了梯度下降技术进行模型的求解。最后，在手写数字识别、人脸识别以及行人重识别三个任务共六个数据集上的实验结果表明，多数情况下本文提出的方法相比于其他度量学习方法取得了更优异的性能。

1　相关工作

为了建立基于数据内在结构特征的度量学习模型，本节首先简要回顾了传统的度量学习，然后介绍了半正定约束度量学习。

1.1　度量学习

度量学习旨在学习一个有效的度量，使同类样本之间的距离能够减小，不同类别样本之间的距离尽可能增大。其中学习基于马氏距离函数的度量学习在计算机视觉领域引起了广泛的兴趣。假定训练样本集合 $\{(\boldsymbol{x}_i, y_i)|i=1,2,\dots,n\}, \boldsymbol{x}_i \in \boldsymbol{R}^m$ 是第 i 个训练样本，y_i 为相应的标签，并且 m 表示样本特征维度个数。则任意两个样本 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 之间的平方马氏距离定义为

$$d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{M}(\boldsymbol{x}_i-\boldsymbol{x}_j) \tag{1}$$

其中：$\boldsymbol{M}\in\boldsymbol{R}^{m\times m}$ 是一个对称半正定矩阵，被称作度量矩阵。从式(1)可以发现在样本给定的情况下，$d_{\boldsymbol{M}}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ 的大小由度量矩阵 $\boldsymbol{M}$ 决定,马氏距离的关键就是学习度量矩阵 $\boldsymbol{M}$。对 $\boldsymbol{M}$ 进行平方根分解，即 $\boldsymbol{M}=\boldsymbol{G}^T\boldsymbol{G}$，其中 $\boldsymbol{G}\in\boldsymbol{R}^{p\times m}(p\le m)$。式(1)可以转换为

$$\begin{aligned} d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j) &= (\boldsymbol{x}_i-\boldsymbol{x}_j)^T\boldsymbol{G}^T\boldsymbol{G}(\boldsymbol{x}_i-\boldsymbol{x}_j) \\ &= [\boldsymbol{G}(\boldsymbol{x}_i-\boldsymbol{x}_j)]^T[\boldsymbol{G}(\boldsymbol{x}_i-\boldsymbol{x}_j)] \\ &= \|\boldsymbol{G}\boldsymbol{x}_i-\boldsymbol{G}\boldsymbol{x}_j\|_2^2 \end{aligned} \tag{2}$$

从式(2)中可以看出，马氏距离实际上寻求一个线性变换，将每个样本 $\boldsymbol{x}_i$ 投影到一个低维子空间 $\boldsymbol{G}\boldsymbol{x}_i$，此低维子空间中任意两个样本的欧式距离等价于原始空间中的马氏距离，因此马氏距离还可以视作标准欧式距离的推广。此外，式(2)可以进一步等价为

$$d_{\boldsymbol{M}}^2(\boldsymbol{x}_i,\boldsymbol{x}_j)=Tr\left(\boldsymbol{M}^T(\boldsymbol{x}_i-\boldsymbol{x}_j)(\boldsymbol{x}_i-\boldsymbol{x}_j)^T\right) \tag{3}$$

其中：$Tr(\cdot)$ 表示矩阵的迹。在接下来的讨论中，本文将构造样本对标签，这里用集合 $\mathcal{S}=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\}$ 表示样本 $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ 有相同的类标签，集合 $\mathcal{D}=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\}$ 表示样本 $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ 有不同的类标签。最终，建立如下相似对标签和不相似对标签：

$$l_{ij}=\begin{cases}1, & \text{if } (\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{D} \\ -1, & \text{if } (\boldsymbol{x}_i,\boldsymbol{x}_j)\in\mathcal{S}\end{cases} \tag{4}$$

1.2　半正定约束度量学习

PCML，作为一种结合了 SVM 的典型度量学习方法，在模式识别和行人重识别等领域展现了优越的性能。PCML 将度量学习问题表述为具有半正定约束的核分类问题，并且通过迭代式训练 SVM 分类器来指导度量矩阵的学习。给定上节中训练样本集，让 $\phi(\boldsymbol{x}_i),\phi(\boldsymbol{x}_j)$ 分别表示样本 $\boldsymbol{x}_i,\boldsymbol{x}_j$ 对应的特征向量，则 PCML 模型的优化问题可以表示为

$$\begin{aligned} &\min_{\boldsymbol{M},b,\xi}\frac{1}{2}\|\boldsymbol{M}\|_F^2+C\sum_{i,j}\xi_{ij} \\ &\text{s.t. } l_{ij}\left(d_{\boldsymbol{M}}^2\left(\phi(\boldsymbol{x}_i),\phi(\boldsymbol{x}_j)\right)-b\right)\ge 1-\xi_{ij},\forall i,j \\ &\xi_{ij}\ge 0, \boldsymbol{M}\succeq 0 \end{aligned} \tag{5}$$

其中：$C>0$ 为误差项的惩罚系数，ξ_{ij} 表示松弛变量，$\|\cdot\|_F$ 表示矩阵的 Frobenius 范数，并且 b 表示距离阈值。此外，注意 $d_{\boldsymbol{M}}(\phi(\boldsymbol{x}_i),\phi(\boldsymbol{x}_j))$ 此时是由特征向量所构建的马氏距离。

2　基于数据内在结构的度量学习

本节首先构建了基于数据驱动的度量学习模型；其次详细介绍了模型的求解过程；最后给出了模型的分类方法。

2.1　模型的构建

PCML 利用 SVM 的大间隔原理来指导马氏距离中度量矩阵的更新，相比于传统没有结合 SVM 的度量学习有着更强的模型泛化能力。然而，SVM 旨在最大化不同类别之间的间隔，只考虑了超平面边界上的样本点，忽略了数据的内在结构信息。事实上，模型的泛化能力不仅与样本类别间的间隔有关，还与数据的内在结构信息有关，尤其是当同类样本之间差异比较大的时候，在这种情况下同类样本间的距离比较大，给 PCML 算法带来了巨大的挑战。

针对以上问题，本文提出了一种基于数据内在结构的度量学习模型。一方面，它不仅利用了 SVM 的大间隔原理，而且还整合了相同类别样本间的类内散度矩阵，这样使得同一类样本间的距离更加近。另一方面，本文还将 L_1-norm 损失函数转换为 L_2-norm 损失函数，这样可以进一步提高模型的泛化性能。

首先针对 PCML 模型，这里先进行一个的转换，让 $\boldsymbol{w}$ 表示矩阵 $\boldsymbol{M}$ 按列组合成的列向量，$\tilde{\boldsymbol{x}}$ 表示矩阵 $(\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j))(\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j))^T$ 按列组合成的列向量，其视作新的特征向量。基于此，可以将 PCML 优化问题转换为

$$\begin{aligned} &\min_{\boldsymbol{M},b,\xi}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+C\sum_{i,j}\xi_{ij} \\ &\text{s.t. } l_{ij}(\boldsymbol{w}^T\tilde{\boldsymbol{x}}-b)\ge 1-\xi_{ij},\xi_{ij}\ge 0,\forall i,j \\ &\boldsymbol{M}\succeq 0 \end{aligned} \tag{6}$$

其中：

$$\begin{aligned} \boldsymbol{w}^T\tilde{\boldsymbol{x}} &= d_{\boldsymbol{M}}^2\left(\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j)\right) \\ &= Tr\left(\boldsymbol{M}^T\left(\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j)\right)\left(\phi(\boldsymbol{x}_i)-\phi(\boldsymbol{x}_j)\right)^T\right) \end{aligned} \tag{7}$$

此时 PCML 模型可以看做一个软间隔 SVM 优化问题，最终本文在此基础上整合了相同类别样本间的类内散度矩阵，并且将 L_1-norm 损失函数转换为 L_2-norm 损失函数，最终建立如下的优化模型：

$$\begin{aligned} &\min_{\boldsymbol{M},b,\xi}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i,j}\xi_{ij}^2 \\ &\text{s.t. } l_{ij}(\boldsymbol{w}^T\tilde{\boldsymbol{x}}-b)\ge 1-\xi_{ij},\xi_{ij}\ge 0,\forall i,j \\ &\boldsymbol{M}\succeq 0 \end{aligned} \tag{8}$$

其中 $\boldsymbol{S}$ 为类内散度矩阵，其定义如下：

$$\boldsymbol{S}=\sum_{\tilde{\boldsymbol{x}}\in\tilde{\boldsymbol{X}}_+}(\tilde{\boldsymbol{x}}-\boldsymbol{m}_c)(\tilde{\boldsymbol{x}}-\boldsymbol{m}_c)^T \tag{9}$$

其中：$\tilde{\boldsymbol{X}}_+$ 表示相似对标签的特征向量组成的集合，$\boldsymbol{m}_c$ 是 $\tilde{\boldsymbol{X}}_+$ 的均值向量，即：

$$\boldsymbol{m}_c=\frac{1}{N}\sum_{\tilde{\boldsymbol{x}}\in\tilde{\boldsymbol{X}}_+}\tilde{\boldsymbol{x}} \tag{10}$$

这里 N 表示 $\tilde{X}_+$ 集合中特征向量的个数。图 1 中概括了本文提出的度量学习模型。首先对输入样本集进行特征提取，得到有效的特征向量，然后构造样本对标签，接下来更新度量矩阵。与 PCML 模型不同的是，在更新度量矩阵的过程中，基于数据内在结构特征的度量学习模型进一步考虑了相同类别样本间的类内散度矩阵，使同一类别间的样本特征靠的更加紧凑，这样使得学习到的度量矩阵鉴别能力更强，从而提高接下来的识别任务。

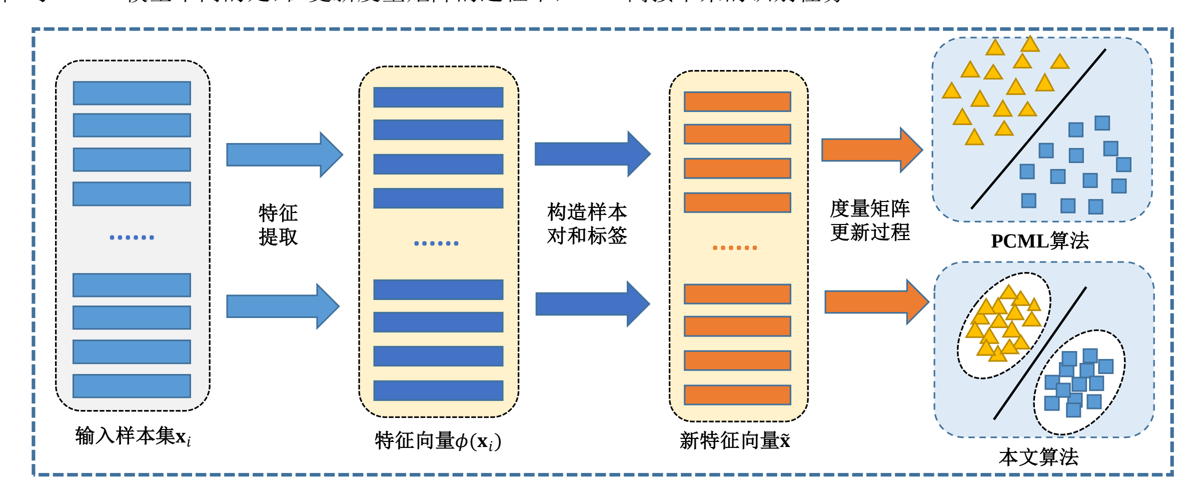


图 1　本文提出的度量学习模型

Fig. 1　The model of this paper proposed metric learning

2.2　模型的求解

为了减少模型的计算复杂度，本文使用了梯度下降技术求解提出的模型式(8)。具体的过程可以分为两个步骤：a) 更新 SVM 中超平面参数 $\boldsymbol{w}$ 和 b；b) 更新度量矩阵 $\boldsymbol{M}$。

1) 更新 SVM 中的参数 $\boldsymbol{w}$ 和 b

对于问题式(8)的求解，可以转变为对偶优化问题进行求解，也可以使用梯度下降技术进行求解。由于使用对偶优化问题求解时间代价比较高，这里为了减少模型的计算开销，本文使用梯度下降技术对问题式(8)进行求解。根据文献[28]，问题式(8)可以转变为一个等价的无约束优化问题：

$$\mathcal{L}(\boldsymbol{w},b)=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i=1}^{2n}\max\left(0,1-l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)\right)^2 \tag{11}$$

其中：$\tilde{\boldsymbol{x}}$ 表示重新组合成的新的特征向量。文献[29]指出，目标函数 $\mathcal{L}(\boldsymbol{w},b)$ 关于参数 $\boldsymbol{w}$ 和 b 是可导的，因此可以通过正常的梯度下降法进行求解。下面分两种情况进行逐一讨论。

首先，当 $l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)<1$ 的时候，此时目标函数 $\mathcal{L}(\boldsymbol{w},b)$ 可以转换如下：

$$\begin{aligned}\mathcal{L}(\boldsymbol{w},b)&=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i=1}^{2n}\|1-l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)\|^2\\&=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i=1}^{2n}(1-l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b))^2\end{aligned} \tag{12}$$

式(12)可以化简为

$$\begin{aligned}\mathcal{L}(\boldsymbol{w},b)&=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i=1}^{2n}\begin{pmatrix}1+(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)^2\\-2l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)\end{pmatrix}\\&=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+C\sum_{i=1}^{2n}\begin{pmatrix}1+\boldsymbol{w}^T\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T\boldsymbol{w}+b^2+\\2b\boldsymbol{w}^T\tilde{\boldsymbol{x}}-2l_i\boldsymbol{w}^T\tilde{\boldsymbol{x}}-2l_ib\end{pmatrix}\end{aligned} \tag{13}$$

式(12)可以进一步化简为

$$\begin{aligned}\mathcal{L}(\boldsymbol{w},b)=&\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}+Cn(1+b^2)+C\boldsymbol{w}^T\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\boldsymbol{w}\\&+2Cb\boldsymbol{w}^T\tilde{\boldsymbol{X}}\boldsymbol{e}-2C\boldsymbol{w}^T\tilde{\boldsymbol{X}}\tilde{\boldsymbol{y}}-2C\boldsymbol{e}^T\tilde{\boldsymbol{y}}b\end{aligned} \tag{14}$$

其中：$\tilde{\boldsymbol{X}}=[\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_{2n}]\in\boldsymbol{R}^{d\times 2n}$ 表示特征向量构建的矩阵，$\tilde{\boldsymbol{y}}=[l_{11},l_{12},\ldots,l_{ij}]\in\boldsymbol{R}^{2n}$ 表示样本对标签组成的向量，并且 $\boldsymbol{e}$ 为单位向量。此时目标函数 $\mathcal{L}(\boldsymbol{w},b)$ 关于参数 $\boldsymbol{w}$ 和 b 的偏导数可以分别表示为

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial\boldsymbol{w}}&=\boldsymbol{S}\boldsymbol{w}+2C\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\boldsymbol{w}+2Cb\tilde{\boldsymbol{X}}\boldsymbol{e}-2C\tilde{\boldsymbol{X}}\tilde{\boldsymbol{y}}\\\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial b}&=2Cnb+2C\boldsymbol{w}^T\tilde{\boldsymbol{X}}\boldsymbol{e}-2C\boldsymbol{e}^T\tilde{\boldsymbol{y}}\end{aligned} \tag{15}$$

其次，当 $l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)\geq 1$ 的时候，目标函数 $\mathcal{L}(\boldsymbol{w},b)$ 可以转换为

$$\mathcal{L}(\boldsymbol{w},b)=\frac{1}{2}\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w} \tag{16}$$

此时目标函数 $\mathcal{L}(\boldsymbol{w},b)$ 关于参数 $\boldsymbol{w}$ 和 b 的偏导数又可以分别表示为

$$\begin{aligned}\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial\boldsymbol{w}}&=\boldsymbol{S}\boldsymbol{w}\\\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial b}&=0\end{aligned} \tag{17}$$

最终，依据上面获得的目标函数关于参数 $\boldsymbol{w}$ 和 b 的偏导数，利用如下的梯度下降算法对参数 $\boldsymbol{w}$ 和 b 进行更新直到收敛：

$$\begin{aligned}\boldsymbol{w}&=\boldsymbol{w}-\mu\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial\boldsymbol{w}}\\b&=b-\mu\frac{\partial\mathcal{L}(\boldsymbol{w},b)}{\partial b}\end{aligned} \tag{18}$$

其中：μ 为步长，也称为学习率。

2) 更新度量矩阵 $\boldsymbol{M}$

为了满足度量矩阵半正定的约束条件，每次利用梯度下降算法更新过 $\boldsymbol{w}$ 和 b 后，还需要利用 $\boldsymbol{w}$ 对度量矩阵 $\boldsymbol{M}$ 进行更新。具体过程如下：首先把 $\boldsymbol{w}$ 重新组合成度量矩阵 $\boldsymbol{M}$，然后为了保证度量矩阵 $\boldsymbol{M}$ 半正定的约束条件，正如文献[26]，本文也在每一次迭代中将度量矩阵 $\boldsymbol{M}$ 向半正定锥上投影的运算。首先对 $\boldsymbol{M}$ 进行特征值分解，转换为如下

$$\boldsymbol{M}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T \tag{19}$$

其中：$\boldsymbol{\Lambda}$ 是度量矩阵 $\boldsymbol{M}$ 的特征值组成的对角矩阵，$\boldsymbol{P}$ 是 $\boldsymbol{M}$ 的正交矩阵。其次执行 $\boldsymbol{\Lambda}_+=\max(\boldsymbol{\Lambda},0)$，最终投影后的度量矩阵为

$$\boldsymbol{M}=\boldsymbol{P}\boldsymbol{\Lambda}_+\boldsymbol{P}^T \tag{20}$$

这里 $\boldsymbol{M}$ 就是最终更新后的度量矩阵，其满足了半正定的约束条件。综合以上模型的求解过程，算法 1 概括了数据驱动的度量学习的模型训练过程：

算法 1　基于数据内在结构特征的度量学习

输入：训练样本集 $\{(\boldsymbol{x}_i,y_i)|i=1,2,\ldots,n\}$，$\boldsymbol{x}_i\in\boldsymbol{R}^m$

输出：最终学习得到的度量矩阵 $\boldsymbol{M}$

1) 初始化 SVM 中的 $\boldsymbol{w}$ 和 b，惩罚系数 C，收敛阈值 σ，最大迭代次数 T，当前迭代次数 $t=0$

2) repeat

3) if $l_i(\boldsymbol{w}^T\tilde{\boldsymbol{x}}+b)<1$：

4) 通过式(15)计算 $\boldsymbol{w}$ 和 b 偏导数

5) else

6) 通过式(17)计算 $\boldsymbol{w}$ 和 b 偏导数

7) 利用计算得到的偏导数通过式(18)更新 SVM 中的 $\boldsymbol{w}$ 和 b；

8) 通过式(20)更新度量矩阵 $\boldsymbol{M}$；

9) $t=t+1$；

10) **until** 满足停止条件(相邻两次迭代的目标函数值的差小于收敛阈值 σ 或者当前迭代次数 t 大于最大迭代次数 T，则迭代过程终止)

11) return 度量矩阵 $\boldsymbol{M}$

2.3 分类方法

一旦获得最终学习的度量矩阵，接下来使用 KNN 算法进行分类。给定一个测试样本，首先应用马氏距离找出该样本点周围最近的 k 个邻近样本点。然后统计该 k 个邻近样本点的标签信息，这些标签中同时属于最多的类别就是测试样本点最终预测的标签。

3 实验设计与分析

为了验证本文所提出的度量学习模型的有效性，本节进行了综合性的实验，第一部分验证了惩罚系数 C 以及 KNN 算法中的 k 对模型性能的影响，同时分析了模型的收敛性。第二部分本文在手写数字识别任务上验证模型的性能。第三部分在人脸识别任务上验证模型的性能。最后在行人重识别的任务上验证模型的性能。此外，本文实验平台为处理器 Intel(R) Core(TM) i7-7700 CPU @3.6GHz 的 64 位 Windows 10 企业版，MATLAB R2016b。

3.1 模型参数对识别率的影响

正如 PCML 模型，本文提出的度量学习模型也仅有 2 个超参数，即惩罚系数 C 以及 KNN 算法中的 k。本文在 PenDigits 手写数字识别数据集[30]上综合调查了参数 C 和 k 对识别准确率的影响，进而选择一个相对比较好的 C 和 k。PenDigits 数据集包含了 0~9 共 10 个类别的 10992 个样本，该数据集里是已经处理过的图像特征共 16 维。实验中选取 7494 个样本作为训练集，剩下的 3498 个样本作为测试集，分别设定参数 C 和 k 的范围为$\{10^{-3},10^{2},10^{-1},10^{0},10^{1},10^{2},10^{3}\}$和 $\{1,2,3,4,5,6,7\}$。此外，为了使模型尽可能收敛到最优的解，本文采用了学习率衰减的策略，设置最大迭代次数为 50，若迭代次数 $t<10$，则学习率 $\mu=0.001$；若 $10\le t<30$，则 $\mu=0.0001$；若 $30\le t<50$，则 $\mu=0.00001$。

图 2(a)报告了模型参数 C 和 k 对识别准确率影响的柱状图。从图中可以明显看出，当惩罚系数 $C=10^{1}$ 和 $k=3$ 的时候，模型的获得了最好的识别准确率。对于惩罚系数 C 的取值，可以发现过小或者过大的惩罚系数都会降低模型的识别准确率。分析其原因，首先，过小的惩罚系数会使得模型过度关注正确识别的样本，使得模型不能正确修正错误识别的样本。其次，过大的惩罚系数使得模型过多的关注被错误识别的样本，迫使模型拟合更复杂的网络参数以尽可能多的修正被错分的训练样本，这无疑增加了模型的优化难度，同时导致模型容易出现过拟合的现象。对于 的取值，过小或者过大的取值同样会降低模型的识别准确率。综合以上分析，在接下来的所有实验中设置参数 $C=4$ 和 $k=3$。

此外，本文也分析了模型的收敛性，图 2(b) 报告了模型的目标函数值随着迭代次数变化的曲线。从图中可以发现，大约经过 30 次的迭代优化之后，模型的目标函数值趋于稳定的状态，结果充分表明了本文所提出度量学习模型是收敛的，这为接下来的实验提供了坚实的基础。

3.2 L1-norm 损失函数和 L2-norm 损失函数结果对比

为了说明本文提出的模型采用 L_2-norm 损失函数替代 L_1-norm 损失函数的有效性，本文也在 PenDigits 手写数字识别数据集上进行了实验验证，实验中的参数 C 和 k 的设定依据 3.1 节中的调查结果，并且结果报告了 10 次平均识别率。

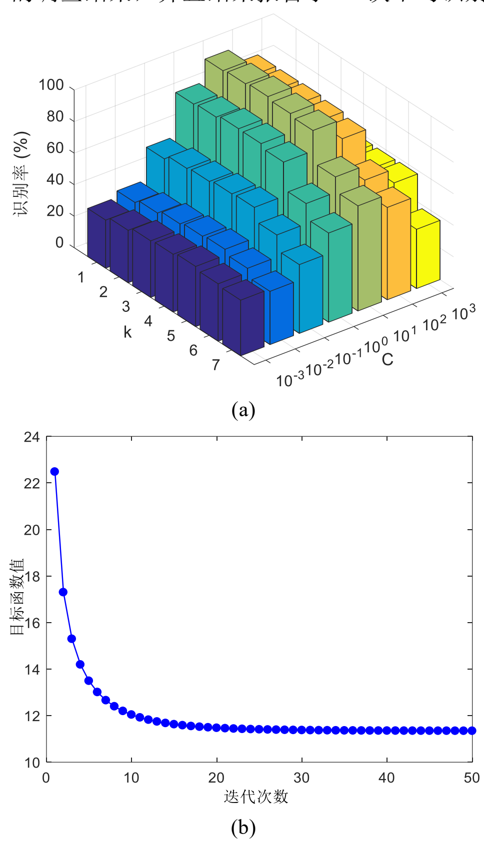


图 2　参数 C 和 k 对识别率的影响以及模型的收敛曲线

Fig. 2　The influence of parameters C and k on the recognition rate and the convergence curve of the proposed model

表 1 列出了模型分别设定为 L_1-norm 损失函数和 L_2-norm 损失函数下的识别率，从中可以发现，相比于模型选择 L_1-norm 损失函数，当模型选择 L_2-norm 损失函数的时候，模型的识别率提升约 0.4%，实验结果表明，本文提出的模型 L_2-norm 损失函数可以进一步提升模型的泛化性能。

表 1　模型在 L_1-norm 损失函数和 L_2-norm 损失函数下的识别率

Tab. 1　The recognition rates of the model on L_1-norm loss function and L_2-norm loss function　/%

损失函数	L_1-norm 损失函数	L_2-norm 损失函数
识别率	98.05	98.42

3.3 手写数字识别

为了验证模型的有效性，本小节在 USPS[31], PenDigits 和 MNIST[32]共 3 个手写数字识别数据集上进行实验分析，这些数据集里全部是包含了 0~9 共 10 个类别的数字。USPS 数据集由 7291 个训练样本和 2007 个测试样本组成，其中每个样本的特征维度是 256。PenDigits 数据集的描述正如 3.1 节。MNIST 数据集包含 60000 个训练样本和 10000 个测试样本，每个样本的特征维度是 784。此外，本文与 8 个经典的度量学习算法进行对比实验，其中包括 ITML[13]，LDML[33]，LMNN[12], DML-eig[34], PLML[35], Doublet-SVM[36], PCML[26] 和 NCML[26]。

在实验中，由于 USPS 和 MNIST 数据集的样本维度比较高，为了提升运算效率，正如文献[26]，本文也使用 PCA 算法将样本特征维度降低到 100。此外实验中的参数 C 和 k 的设定依据 3.1 节中的调查结果，同时对数据集中的所有样本进行了归一化处理。表 2 列出了本文提出的算法以及其他的算法在 3 个手写数字识别数据集上运行 10 次的平均识别准确率，从中表中可以发现，本文提出的算法取得了最好的

识别准确率。与 PCML 算法相比，本文算法提升了约 0.6%的识别准确率，这充分说明了简单的考虑不同类别样本间的最大间隔限制了模型识别准确率，通过整合同类样本间的类内散度矩阵，使同类样本可以进一步提升模型的识别准确率。

表 2　不同算法在 3 个手写数字识别数据集上的识别率

Tab. 2　The recognition rates of different algorithms on three handwritten digit datasets　/%

算法	USPS	PenDigits	MNIST
ITML	93.68	97.70	97.12
LDML	94.88	97.68	93.96
LMNN	94.66	97.77	97.72
DML-eig	94.56	96.30	94.92
PLML	93.25	97.57	97.48
Doublet-SVM	94.57	97.57	96.81
PCML	94.68	97.85	96.25
NCML	94.70	97.87	97.45
本文算法	**95.11**	**98.33**	**97.98**

此外，图 3 展示了本文提出的算法和其他算法分别在 3 个手写数字识别数据集上的训练时间，从图中可以发现，本文提出算法在 USPS 和 PenDigits 数据集上的训练时间仅仅高于 Double-SVM，在 MNIST 数据集上了本文提出的算法取得了最短的训练时间。重要的是，本文提出的算法在所有的数据集上训练时间明显低于 PCML 算法，分析其原因，在求解度量矩阵时，本文提出的方法直接使用了梯度下降方法来优化度量矩阵，而 PCML 把原问题转换成了对偶问题，通过求解对偶问题来获得度量矩阵，这样增加了时间消耗。实验结果进一步表明本文提出的方法可以提升模型的计算效率。

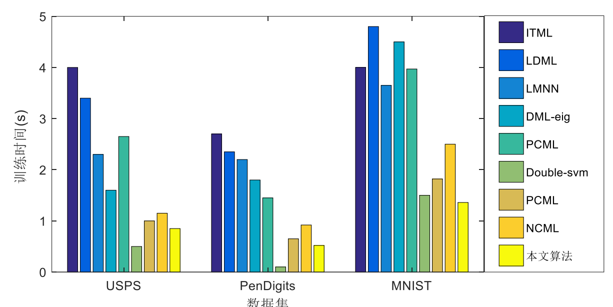


图 3　不同算法在 3 个手写数字识别数据集上的训练时间

Fig. 3　The training time of different algorithms on three handwritten digit datasets

3.4　人脸识别

为了进一步验证模型的有效性，本小节在大型人脸识别数据集 LFW[37]上进一步实验分析。LFW 数据集是人脸识别常用的数据集，它包含了 5749 人的 13233 张带标签的人脸面部，其中 1680 人有两张或更多的面孔。而且，由于面部在尺度、姿势、光线、背景、表情、发型和眼镜上的变化，同时这些人脸面部是在野外场景中拍摄，致使该数据集非常具有挑战性。此外，该数据集提供了 5400 对人脸图像进行训练，300 对正、300 对负进行测试。

在实验中，为了获取人脸图像有效的特征，参考文献[24]的处理方式，本文也使用人脸图像的 VGG-Face 特征[38]来评估算法的识别准确率，同时由于 VGG-Face 特征维度比较高(4096)，本文采用 PCA 算法把特征维度降低到 50。此外，实验中的参数 C 和 k 的设定依据 3.1 节中的调查结果，对于每个算法本文统计了 10 次识别准确率，并且报告了平均识别准确率。这里与 8 个经典的算法进行了对比实验，其中包括 ITML，DML-eig， KISSME[14]，XQDA[39]，Doublet-SVM，PCML 和 NCML。

表 3 列出了不同算法在 LFW 数据集上的识别准确率以及训练时间。从表中可以看出，相比其他算法，本文提出的算法获得了最好的识别率，相比于与 PCML 和 NCML 提高了约 1%，这说明本文提出模型在同时考虑不同类别样本之间的最大间隔和相同类别样本之间的类内散度的情况下，可以进一步提升度量矩阵的鉴别能力，从而提升模型的识别率。此外，从表中还可以发现，本文提出的算法的训练时间明显少于 ITML，DML-eig，PCML 和 NCML，多于 KISSME，XQDA 和 Doublet-SVM，其中 Doublet-SVM 和 KISSME 是一次优化方法，XQDA 是一个子空间方法，本文提出的方法使用了梯度下降技术直接对原问题进行了优化，所以它们可以获得更短的训练时间。以上分析再次论证了本文提出的算法相比于 PCML 算法不仅提升了识别准确率，而且可以进一步提升模型的计算效率。

表 3　不同算法在 LFW 数据集上的识别率

Tab. 3　The recognition rates of different algorithms on LFW dataset　/%

算法	识别率	训练时间
ITML (VGG-Face)	96.37	194.92
DML-eig(VGG-Face)	94.92	256.24
KISSME(VGG-Face)	96.40	**0.05**
XQDA(VGG-Face)	95.66	0.10
Doublet-SVM(VGG-Face)	96.37	0.39
PCML(VGG-Face)	96.41	9.15
NCML(VGG-Face)	96.42	9.88
本文算法(VGG-Face)	**96.87**	7.02

3.5　行人重识别

最后，本小节在 CUHK01[40]和 CUHK03[41]共 2 个行人重识别数据集上进一步验证模型的有效性。CUHK01 数据集共包含 971 位行人的 3,884 张图像，这些图像是由 CUHK 校园里的 2 台摄像机拍摄的，其中每个行人在每个摄像头下面平均有 2 张图像。CUHK03 数据集共包含 1476 位行人的 14096 幅图像，这些图像同样是由 CUHK 校园里的两台摄像机拍摄的，其中每位行人在每个摄像头下面平均有 4~8 张图像。另外，CUHK03 提供了 labeled 和 detected 两个数据集，本文对这两个数据集分别进行了实验。

在实验中，对于 CUHK01 数据集，正如 PCML，训练集包含了 485 人的图像，这些图像是从这个数据集中随机选取的，剩下的 486 人的图像构成了测试集。对于 CUHK03 数据集，参考文献[24]的处理方式，本文随机选择 1367 人的图像作为训练集,并使用其余 100 人的图像作为测试集。此外实验中的参数 C 和 k 的设定依据 3.1 节中的调查结果，同时对数据集中的所有样本进行了归一化处理。此外对于所有的算法，为了获取有效的特征表示来提升模型的识别准确率，参考文献[26]处理方式，本文也提取了行人图像的 LOMO 特征[42]。实验中本文与 8 个具有代表性的度量学习算法进行了对比实验，其中包括 LMNN，LDML，DML-eig，KISSME，XQDA，PCML，Doublet-SVM 和 NCML。

表 4~6 报告了本文提出的算法以及其他的算法在 2 个行人重识别数据集上运行 10 次的平均识别准确率。从表中还可以发现，本文提出的算法获得了最好的模型识别率。从表 3 可以知道，相比于 PCML，本文算法的识别率上提高了约 1.5%。从表 4~5 中可以知道，相比于 PCML，本文提出的算法的识别率提高了约 1.2%。此外，从表 4~6 中可以进一步发现，在训练时间上，本文提出的算法多数情况下获得了最短的训练时间，仅仅在 CUHK01 数据集上的训练时间高于 XQDA 算法。然而重要的是，相比于 PCML，本文提出的方法训练时间缩短了很多，例如表 4 中本文提出的算法的训练时间 156.3 秒，PCML 的训练时间 482.65 秒，其中 PCML 通过求解对偶问题来优化度量矩阵，而本文提出的算法直接对

原问题利用梯度下降法进行优化，这样避免了求解对偶问题带来了更多的时间开销。以上的实验结果再一次说明本文提出算法通过整合不同类别样本之间的最大间隔和相同类别样本之间的类内散度矩阵，可以使得学习到的度量矩阵有更强的鉴别能力，进而提升模型的识别准确率，而且还可以进一步提升模型的计算效率。

表 4 不同算法在 CUHK01 数据集上的识别率以及训练时间

Tab. 4 The recognition rates and the training time of different algorithms on CUHK01 dataset

算法	Rank-1/%	Rank-5/%	Rank-10/%	Rank-20/%	训练时间/s
LMNN(LOMO)	55.28	83.45	89.21	97.86	2558.60
LDML(LOMO)	51.36	77.69	87.35	96.24	63.42
DML-eig(LOMO)	51.31	81.96	91.56	98.04	3330.8
ITML (LOMO)	57.38	82.41	90.24	97.16	4342.31
XQDA(LOMO)	63.18	85.04	91.87	97.68	**15.47**
Doublet-SVM(LOMO)	51.58	76.56	86.23	93.46	22.18
PCML(LOMO)	61.56	84.02	90.32	98.07	33.31
NCML(LOMO)	61.73	84.43	90.56	98.19	58.30
本文算法(LOMO)	**63.68**	**85.24**	**91.96**	**98.96**	16.83

表 5 不同算法在 CUHK03 中的 labeled 训练集上的识别率以及训练时间

Tab. 5 The recognition rates and the training time of different algorithms on labeled training set in CUHK03

算法	Rank-1/%	Rank-5/%	Rank-10/%	Rank-20/%	训练时间/s
LMNN(LOMO)	50.96	80.53	89.01	94.96	7896.25
LDML(LOMO)	51.18	81.96	89.85	95.24	732.42
DML-eig(LOMO)	17.96	50.26	66.78	84.35	496.75
ITML (LOMO)	46.86	80.06	88.24	90.28	1135.27
XQDA(LOMO)	52.23	83.42	90.56	95.84	862.3
Doublet-SVM(LOMO)	51.35	82.14	90.08	95.64	198.36
PCML(LOMO)	53.26	84.05	91.02	96.85	482.65
NCML(LOMO)	53.45	84.36	91.24	97.01	584.28
本文算法(LOMO)	**54.18**	**85.28**	**91.56**	**97.31**	**156.3**

表 6 不同算法在 CUHK03 中的 detected 训练集上的识别率以及训练时间

Tab. 6 The recognition rates and the training time of different algorithms on detected training set in CUHK03

算法	Rank-1/%	Rank-5/%	Rank-10/%	Rank-20/%	训练时间/s
LMNN(LOMO)	44.58	77.85	85.84	88.09	7896.25
LDML(LOMO)	45.36	78.29	86.56	89.13	732.42
DML-eig(LOMO)	14.08	42.68	60.58	82.36	496.75
ITML (LOMO)	44.16	78.06	86.24	90.11	1135.27
XQDA(LOMO)	46.18	80.28	87.58	90.56	862.3
Doublet-SVM(LOMO)	45.24	79.25	86.76	89.25	198.36
PCML(LOMO)	46.18	80.09	87.24	89.78	482.65
NCML(LOMO)	46.88	80.35	87.57	90.12	584.28
本文算法(LOMO)	**47.83**	**81.23**	**88.88**	**90.78**	**156.3**

4 结束语

PCML 算法仅仅考虑了不同类别样本之间的最大间隔，忽略了每一次迭代过程中同一类别间的样本特征空间的也在发生变化，这样限制了模型泛化能力。基于此，本文提出了一种基于数据内在结构特征的度量学习算法，该算法不仅考虑不同类别样本之间的最大间隔，而且相同类别样本之间的类内散度。在手写数字识别，人脸识别，行人重识别三个任务上的实验结果表明，本文提出的算法最终学习到的度量矩阵有更强的鉴别能力，取得了更高的识别准确率。

然而，虽然本文所提出的算法取得了较好的性能，但仍存在一些问题。一方面，正如 PCML，本文提出的算法也需要手动的设定惩罚系数和近邻参数。另一方面，本文提出的算法利用了迭代优化策略去求解优化问题，这并不能保证求得的解是全局最优解。因此，在未来的工作中，本文将进一步讨论如何解决这些问题。

参考文献：

[1] Yang Weiming, Liu Yuliang, Zhang Quan *et al.* Comparative Object Similarity Learning-Based Robust Visual Tracking [J]. IEEE Access, 2019: 50466-50475.

[2] Zhang Shengping, Qi Yuankai, Jiang Feng *et al.* Point-to-Set Distance Metric Learning on Deep Representations for Visual Tracking [J]. IEEE Transactions on Intelligent Transportation Systems, 2017: 1-12.

[3] Ramos J, Kockelkorn T, Ramos I, *et al.* Content-Based Image Retrieval by Metric Learning From Radiology Reports: Application to Interstitial Lung Diseases [J]. Biomedical and Health Informatics, 2016, 20 (1): 281-292.

[4] Yu Jun, Yang Xiaokang, Gao Fei, *et al.* Deep Multimodal Distance Metric Learning Using Click Constraints for Image Ranking [J]. IEEE Transactions on Cybernetics, 2016: 1-11.

[5] Guillaumin, Matthieu and Verbeek *et al.* Multiple instance metric learning from automatically labeled bags of faces [C]// Proc of European Conference on Computer Vision, 2010: 634-647.

[6] Cai Lipeng, Ying Shihui, Peng Yaxin, *et al.* Intrinsic Metric Learning with Subspace Representation [J]. IEEE Access, 2019: 1-1.

[7] Bac N, Ferri F J, Carlos M, *et al.* An efficient method for clustered multi-metric learning [J]. Information Sciences, 2018: S002002551830673X-.

[8] Duan Yueqi, Zheng Wenzhao, Lin Xudong, *et al.* Deep Adversarial Metric Learning [J]. IEEE Transactions on Image Processing, 2020: 2037-2051.
[9] Chu Ruihang, Sun Yifan, Li Yadong, *et al.* Vehicle Re-identification with Viewpoint-aware Metric Learning [C]// Proc of IEEE International Conference on Computer Vision, 2019.
[10] Hu Haimiao, Fang Wen, Li Bo, *et al.* An Adaptive Multi-Projection Metric Learning for Person Re-identification across Non-Overlapping Cameras [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2018: 1-1.
[11] Goldberger J, Roweis S, Hinton G, *et al.* Neighborhood components analysis [C]// Proc of International Conference on Neural Information Processing Systems, 2004.
[12] Weinberger K Q. Distance Metric Learning for Large Margin Nearest Neighbor Classification [J]. Journal of Machine Learning Research, 2009.
[13] Jason V. Davis, Brian Kulis, Prateek Jain, *et al.* Information-theoretic metric learning [C]// Proc of the 24th International Conference on Machine Learning, 2007: 20-24.
[14] Kstinger M, Hirzer M, Wohlhart P, *et al.* Large Scale Metric Learning from Equivalence Constraints [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[15] Belkin, Mikhail, Niyogi, *et al.* Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples [J]. Journal of Machine Learning Research, 2006: 2399-2434.
[16] Zhang Qin, Tian Yingjie, Liu Dalian. Nonparallel Support Vector Machines for Multiple-Instance Learning [J]. Procedia Computer Science, 2013, 17: 1063-1072.
[17] Sain, Stephan R. The Nature of Statistical Learning Theory [J]. Techno metrics, 1996, 38 (4): 409-409.
[18] Chih-Chung, Chang, Chih-Jen, *et al.* LIBSVM: A library for support vector machines [J]. ACM Transactions on Intelligent Systems & Technology, 2011.
[19] Tsang, Ivor W, Kwok, James T, Cheung, PakMing. Core vector machines: fast SVM training on very large data sets [J]. Machine Learning Research, 2005, 6 (2): 363-392.
[20] A. Bordes, L. Bottou, P. Gallinari, *et al.* Solving multiclass support vector machines with LaRank [C]// Proc of the 24th International Conference on Machine Learning, 2007: 89–96.
[21] Teo C H, Le Q, Smola A, *et al.* A Scalable Modular Convex Solver for Regularized Risk Minimization [C]// Proc of ACM Sigkdd International Conference on Knowledge Discovery & Data Mining, 2007.
[22] Shalev-Shwartz S, Singer Y, Srebro N, *et al.* Pegasos: Primal estimated sub-GrAdient sOlver for SVM [C]// Proc of the 24th International Conference on Machine Learning, 2007.
[23] Nam Nguyen, Guo Yunsong. Metric Learning: A Support Vector Approach [C]// Proc of European Conference, 2008.
[24] Brunner C, Fischer A, Luig K, *et al.* Pairwise Support Vector Machines and their Application to Large Scale Problems [J]. Journal of Machine Learning Research, 2012, 13 (1): 2279-2292.
[25] H. Do, A. Kalousis, J. Wang, *et al.* A metric learning perspective of SVM: On the relation of SVM and LMNN. [Online]. Available: https://arxiv. org/abs/1201. 4714.
[26] Zuo Wangmeng, Wang Faqiang, Zhang Da, *et al.* Distance Metric Learning via Iterated Support Vector Machines [J]. IEEE Transactions on Image Processing, 2017: 1-1.
[27] Chung K M, Kao W C, Sun C L, *et al.* Radius Margin Bounds for Support Vector Machines with the RBF Kernel [J]. Neural Computation, 2002, 15: 2643-2681.
[28] Lin Liang, Wang Keze, Zuo Wangmeng, *et al.* A Deep Structured Model with Radius-Margin Bound for 3D Human Activity Recognition [J]. 2015.
[29] Chapelle O. Training a Support Vector Machine in the Primal [J]. Neural computation, 2007, 19 (5): 1155-1178.
[30] Zheng Yi, Zhao Li, Sun Jie, *et al.* High performance quadratic classifier and the application on pendigits recognition [J]. Proceedings of the IEEE Conference on Decision & Control, 2007: 3072-3077.
[31] Ruppert, David. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [J]. Journal of the American Statistical Association, 2004, 99 (466): 567-567.
[32] Lecun Y, Bottou L. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
[33] M. Guillaumin, J. Verbeek and C. Schmid, "Is that you?Metric learning approaches for face identification [C]// Proc of the 12th IEEE International Conference on Computer Vision, 2009: 498-505.
[34] Ying Yiming, Li Ping. Distance Metric Learning with Eigenvalue Optimization [J]. Journal of Machine Learning Research, 2012, 13 (1): 1-26.
[35] Wang J, Woznica A, Kalousis A. Parametric Local Metric Learning for Nearest Neighbor Classification [J]. Advances in neural information processing systems, 2012: 1601-1609.
[36] Wang Faqing, Zuo Wangmeng, Zhang Da, *et al.* A Kernel Classification Framework for Metric Learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26 (9): 1950-1962.
[37] G. B. Huang, M. Ramesh, T. Berg, *et al.* Labeled faces in the wild: A database for studying face recognition in unconstrained environments [J]. Massachusetts, 2007: 07-49.
[38] O. M. Parkhi, A. Vedaldi, *et al.* Deep face recognition [C]// Proc of British Machine Vision Conference, 2015: 41–52.
[39] Song Mengye, Gong Shengrong, Liu Chunping *et al.* Person re-identification by improved Local Maximal Occurrence with color names [C]// Proc of International Congress on Image & Signal Processing. IEEE, 2015.
[40] Li Wei, Zhao Rui, Wang Xiaogang. Human Reidentification with Transferred Metric Learning [C]// Proc of the 11th Asian conference on Computer Vision-Volume Part I. Springer, Berlin, Heidelberg, 2012.
[41] Wei Li, Zhao Rui, Xiao Tong, *et al.* DeepReID: Deep Filter Pairing Neural Network for Person Re-identification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014.
[42] Liao Shengcai, Hu Yang, Zhu Xiangyu, *et al.* Person re-identification by local maximal occurrence representation and metric learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2015.